



ACADEMIC
PRESS

Biochemical and Biophysical Research Communications 294 (2002) 949–955

BBRC

www.academicpress.com

The P1' specificity of tobacco etch virus protease

Rachel B. Kapust,^a József Tözsér,^b Terry D. Copeland,^a and David S. Waugh^{a,*}

^a *Macromolecular Crystallography Laboratory, Center for Cancer Research, National Cancer Institute at Frederick, P.O. Box B, Frederick, MD 21702, USA*

^b *Department of Biochemistry and Molecular Biology, Medical and Health Science Center, University of Debrecen, Debrecen, Nagyerdei krt. 98 H-4012, Hungary*

Received 14 May 2002

Abstract

Affinity tags have become indispensable tools for protein expression and purification. Yet, because they have the potential to interfere with structural and functional studies, it is usually desirable to remove them from the target protein. The stringent sequence specificity of the tobacco etch virus (TEV) protease has made it a useful reagent for this purpose. However, a potential limitation of TEV protease is that it is believed to require a Gly or Ser residue in the P1' position of its substrates to process them with reasonable efficiency. Consequently, after an N-terminal affinity tag is removed by TEV protease, the target protein will usually retain a non-native Ser or Gly residue on its N-terminus, and in some cases this may affect its biological activity. To investigate the stringency of the requirement for Gly or Ser in the P1' position of a TEV protease recognition site, we constructed 20 variants of a fusion protein substrate with an otherwise optimal recognition site, each containing a different amino acid in the P1' position. The efficiency with which these fusion proteins were processed by TEV protease was compared both in vivo and in vitro. Additionally, the kinetic parameters K_M and k_{cat} were determined for a representative set of peptide substrates with amino acid substitutions in the P1' position. The results indicate that many side-chains can be accommodated in the P1' position of a TEV protease recognition site with little impact on the efficiency of processing. © 2002 Published by Elsevier Science (USA).

Keywords: TEV protease; Fusion protein; Affinity tag; N-end rule

In the field of proteomics, one of the main challenges is developing efficient methods for high-throughput expression and purification of soluble proteins for structural and functional studies. High-throughput purification requires a general approach that will be effective for proteins with diverse chemical properties. Genetically engineered affinity tags are ideal for this purpose because they can be exploited to devise generic purification protocols. Additional benefits of affinity tags such as *Escherichia coli* maltose-binding protein (MBP) [1] and *Schistosoma japonicum* glutathione S-transferase [2] include the ability to protect their fusion partners (passenger proteins) from intracellular proteolysis [3,4] and increase their yield [5]. A further advantage of MBP is its remarkable ability to enhance the solubility of its passenger proteins [4,6,7].

Because of concerns about the impact of affinity tags on the structure or activity of a passenger protein, it is ordinarily desirable to remove them. It is this step in the process that has proven to be the Achilles heel of the fusion approach. Both chemical and enzymatic methods have been used to cleave fusion proteins at engineered sites [5,8,9]. However, chemical reagents suffer from a lack of specificity and often work effectively only under severe conditions that may irreversibly damage the passenger protein [10]. Enzymatic reagents, on the other hand, function under milder reaction conditions and typically exhibit greater sequence specificity.

Most fusion proteins are engineered so that the N-terminus of the passenger protein is genetically fused to the C-terminus of the affinity tag with a linker region containing a protease recognition site between them. In principal, enteropeptidase (enterokinase) and factor Xa can be used to generate passenger proteins with native N-termini after digestion of a fusion protein substrate because their primary specificity determinants are

* Corresponding author. Fax: +1-301-846-7148.

E-mail address: waughd@ncicrf.gov (D.S. Waugh).

N-terminal to the scissile bond. In practice, however, both of these proteases have been observed to cleave fusion proteins at locations other than the intended target site, often resulting in the degradation of the passenger protein [10–13].

In contrast to the aforementioned proteases, there have been no reported examples of cleavage at non-canonical sites in fusion proteins by the nuclear inclusion protease from tobacco etch virus (TEV protease). A potential disadvantage of TEV protease, however, is that its S1' pocket is presumed to be an important specificity determinant; amino acids other than Gly or Ser in this position are alleged to reduce proteolytic efficiency [14,15]. To investigate how stringent the requirement is for a Gly or Ser residue in the P1' position of the TEV protease cleavage site, we used a model fusion protein substrate, MBP–NusG, containing a canonical recognition site (ENLYFQ/G) that is processed very efficiently by TEV protease. Twenty MBP–NusG fusion proteins, each with a different amino acid in the P1' position, were constructed by site-directed mutagenesis. The efficiency with which these P1' variants were cleaved by TEV protease was compared both *in vivo*, using an intracellular processing system, and *in vitro*. Additionally, the kinetic parameters K_M and k_{cat} were determined for a representative set of peptide substrates with amino acid substitutions in the P1' position. These experiments constitute the first comprehensive investigation of P1' specificity in a potyviral protease.

Materials and methods

Plasmid expression vectors. pRK603 and bacterial strain DH5 α Z1 were described previously [16]. pKM631, which produces an MBP–NusG fusion protein with Gly in the P1' position of the TEV protease recognition site, was constructed by polymerase chain reaction (PCR) amplification of the open reading frame (ORF) encoding NusG from *Aquifex aeolicus* genomic DNA with the following oligonucleotide primers: 5'-GAG AAC CTG TAC TTC CAG GGT ATG AGC GAG CAA CAG GTT CAG GAA C-3' and 5'-ATT AGT GAT GAT GGT GGT GAT GAA TCT TTT CCA CTT GGT CAA AGT CCA G-3'. This PCR amplicon was subsequently used as the template for another PCR with primers PE-277 and PE-278 [17], generating a second amplicon that was inserted by recombinational cloning into pKM596 [18] to yield pKM631.

The oligodeoxyribonucleotides used to construct plasmid expression vectors for the production of MBP–NusG fusion proteins with altered TEV protease recognition sites were: PE-431 (forward outer PCR primer), 5'-GGT TAA TAA AGA CAA ACC GCT GGG TG-3'; PE-430 (reverse outer PCR primer), 5'-GCC ATG AGG AGC TTG TCG TTC ATG TG-3'; PE-X (forward mutagenic PCR primer), 5'-TAC TTC CAG XXX ATG AGC GAG CAA CAG G-3'; PE-Y (reverse mutagenic PCR primer), 5'-GTT GCT CGC TCA TXX XCT GGA AGT ACA GG-3'. The MBP–NusG expression vectors were constructed by overlap extension PCR [19]. First, two different PCR reactions were performed using pKM631 as the template: one with PE-431 and the appropriate reverse mutagenic primer (PE-Y), and the other with PE-430 and the corresponding forward mutagenic primer (PE-X). These two PCR amplicons were then combined and used as

the template for a third PCR reaction with the outer primers, PE-431 and PE-430. The final PCR product was digested with *Nco*I and *Pst*I, and then ligated with the *Nco*I/*Pst*I vector fragment of pKM631. The same procedure was used to construct all of the P1' variants. The nucleotide sequence of each construct was confirmed experimentally.

***In vivo* processing experiments.** Cells from single, drug-resistant colonies of *E. coli* DH5 α Z1 containing an MBP–NusG fusion protein vector and pRK603, the TEV protease expression vector, were grown to saturation in 5 ml Luria broth [20] supplemented with the appropriate antibiotics (100 μ g/ml ampicillin and 30 μ g/ml kanamycin) at 37 °C. The saturated culture was diluted 1:50 in the same medium and grown to early log phase ($A_{600} = 0.3$ to 0.5) at 37 °C, at which point the temperature was shifted to 30 °C (the optimum temperature for TEV protease processing) and both isopropyl- β -D-thiogalactopyranoside (IPTG) and anhydrotetracycline (final concentrations of 1 mM and 100 ng/ml, respectively) were added to initiate the production of the fusion protein and TEV protease. After 4 h of shaking at 30 °C, the cells from 10 ml of each culture were recovered by centrifugation and resuspended in 1 ml of 20 mM Tris–HCl (pH 8.0), 1 mM EDTA. The cell suspensions were lysed by sonication, after which aliquots of the cell lysate were mixed with an equal volume of 2 \times SDS sample buffer [21] to produce samples of the total intracellular protein for sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE). The samples were heated at 90 °C for 4 min and then centrifuged at 14,000g for 15 min prior to SDS–PAGE. Samples were analyzed on 10–20% Tris–glycine SDS–polyacrylamide gels (Novex) and visualized by staining with Gelcode Blue (Pierce).

Overproduction and purification of MBP–NusG fusion proteins. DH5 α Z1 cells containing an MBP–NusG fusion protein expression vector were grown to saturation at 37 °C in Luria broth supplemented with 100 μ g/ml ampicillin. The saturated culture was diluted 1:50 into 1 L of the same medium and grown in a shake-flask to early log phase ($A_{600} = 0.3$ to 0.5). At this point, IPTG was added to a final concentration of 1 mM, and the culture was grown for an additional 4 h at 37 °C. The culture was centrifuged at 10,000g for 10 min, and the cell pellet was stored at –80 °C.

The cell pellet was dissolved in 10 ml lysis buffer (100 mM Na₂HPO₄, pH 7.8, 200 mM NaCl, 5 mM β -mercaptoethanol (BME), 0.1% Tween 20). The cells were lysed by sonication and then centrifuged at 10,000g for 20 min. The supernatant was mixed with 4 ml Ni–NTA resin (Qiagen), previously equilibrated with cell lysis buffer, and gently rocked at 4 °C. After 1 h, the resin was loaded into a BioRad Poly-Prep column and then washed with 10 ml lysis buffer. The column was then washed with 20 ml of 100 mM Na₂HPO₄, pH 7.4, 200 mM NaCl, 5 mM BME, 10% glycerol, and 20 mM imidazole. Finally the MBP–NusG fusion protein was eluted from the column with 4 ml of 100 mM Na₂HPO₄, pH 7.4, 200 mM NaCl, 5 mM BME, 5% glycerol, and 300 mM imidazole.

To concentrate the protein for dialysis, 8 ml of saturated ammonium sulfate solution was added to the eluate, and the precipitate was recovered by centrifugation at 7000g for 20 min. The pellet was resuspended in 2 ml of TEV protease reaction buffer (50 mM Tris–HCl, pH 8.0, 0.5 mM EDTA, 1 mM DTT) and then dialyzed extensively against the same buffer at 4 °C.

***In vitro* processing experiments.** The TEV protease used for the *in vitro* processing experiments has an amino acid substitution (S219V) that greatly reduces its rate of autolysis without compromising its catalytic activity [22]. All reactions were performed in TEV protease reaction buffer at 30 °C. The concentration of the MBP–NusG fusion proteins was held constant at 1 mg/ml (14 μ M), while the concentration of TEV protease varied depending on the efficiency with which each P1' variant was processed. TEV protease concentrations were either 0.005 mg/ml (0.18 μ M), 0.013 mg/ml (0.47 μ M), 0.04 mg/ml (1.4 μ M), or 0.2 mg/ml (7 μ M). The *in vitro* processing reactions were initiated by combining the enzyme and substrate in 200 μ l of TEV protease reaction buffer. At 30 min intervals, 15 μ l aliquots was removed from the reaction over the course of 3.5 h and mixed with 85 μ l 1 \times SDS sample

buffer. The samples were denatured at 90 °C for 4 min and then subjected to SDS-PAGE. The proteins were visualized by staining with Gelcode Blue.

Statistical analysis and densitometry. At least three independent experiments were performed to obtain numerical estimates of the fraction of each fusion protein that was processed by TEV protease *in vitro*. Gelcode Blue stained gels were scanned with a Molecular Dynamics Personal Densitometer, and then the pixel densities of the bands corresponding to the intact fusion protein and the individual MBP and NusG moieties were obtained directly by volumetric integration. Using these pixel density values, the percentage of processed fusion protein at an individual time point was calculated by dividing the sum of the values for the processed MBP and NusG by the total amount of intact fusion protein plus the previously calculated MBP and NusG sum. Using Microsoft Excel, regression lines were drawn and error bars were calculated based on the standard deviation of the population.

Oligopeptide synthesis and characterization. Oligopeptide substrates for TEV protease (Thr–Glu–Asn–Leu–Tyr–Phe–Gln–Xaa–Gly–Thr–Arg–Arg–NH₂) were synthesized by standard 9-fluorenylmethyloxycarbonyl chemistry on a model 430A automated peptide synthesizer (Applied Biosystems). Stock solutions were made in distilled water and the peptide concentrations were determined by amino acid analysis, following hydrolysis in 6N HCl, using a Beckman 6300 amino acid analyzer.

Enzyme kinetics. TEV protease assays were initiated by mixing 20 μ l of TEV (S219V) protease solution (40–1500 nM) in 50 mM Na-phosphate, pH 7.0, 5 mM DTT, 800 mM NaCl, and 10% glycerol, with 20 μ l of substrate solution (0.04–5.2 mM). The actual range of substrate concentrations was selected on the basis of the approximate K_M values. Measurements were performed at six different substrate concentrations. The reaction mixtures were incubated at 30 °C for 30 min and then stopped by the addition of 160 μ l 4.5 M guanidine–HCl containing 1% trifluoroacetic acid. An aliquot was injected onto a Nova-Pak C₁₈ reverse-phase chromatography column (3.9 \times 150 mm, Waters Associates) using an automatic injector. The substrates and reaction products were separated by an increasing water–acetonitrile gradient (0–100%) in the presence of 0.05% TFA. To determine the correlation between peak areas of the cleavage products and their amount, fractions were collected and subjected to amino acid analysis. The k_{cat} values were calculated by assuming 100% activity for the enzyme. Kinetic parameters were determined by fitting the data obtained at less than 20% substrate hydrolysis to the Michaelis–Menten equation, using the Fig. P program (Fig. P Software Corp.). Standard deviations for the k_{cat}/K_M values were calculated as described [23]. The k_{cat}/K_M value for the peptide with Val in the P1' position was determined from the linear part of the rate versus concentration profile.

Results

Construction of MBP–NusG fusion proteins

A bipartite fusion between *E. coli* MBP and *Aquifex aeolicus* NusG (MBP–NusG) was used as a model substrate to investigate the P1' specificity of TEV protease (Fig. 1). This fusion protein was selected primarily because it is an efficient substrate for TEV protease, but also because it is well expressed and highly soluble in *E. coli*. Twenty MBP–NusG fusion protein expression vectors were assembled by overlap-extension PCR [19], each one encoding a different amino acid in the P1' position of an otherwise canonical TEV protease recognition site (ENLYFQX). Hexahistidine tags were

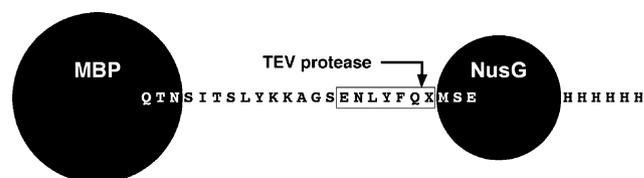


Fig. 1. Schematic representation of MBP–NusG fusion protein substrates for TEV protease. The P1' residue in the TEV protease cleavage site is designated “X.”

added to the C-termini of the MBP–NusG fusion proteins to facilitate their purification.

Intracellular processing of fusion proteins

Initially, we exploited an intracellular processing system [16] to compare the efficiency with which the twenty P1' variants were cleaved by TEV protease *in vivo*. TEV protease was co-expressed with each of the MBP–NusG fusion proteins in *E. coli* DH5 α Z1 cells, and the composition of the total intracellular protein was examined by SDS-PAGE (Fig. 2). Remarkably, every fusion protein except the Pro variant was processed to some extent under these conditions, suggesting

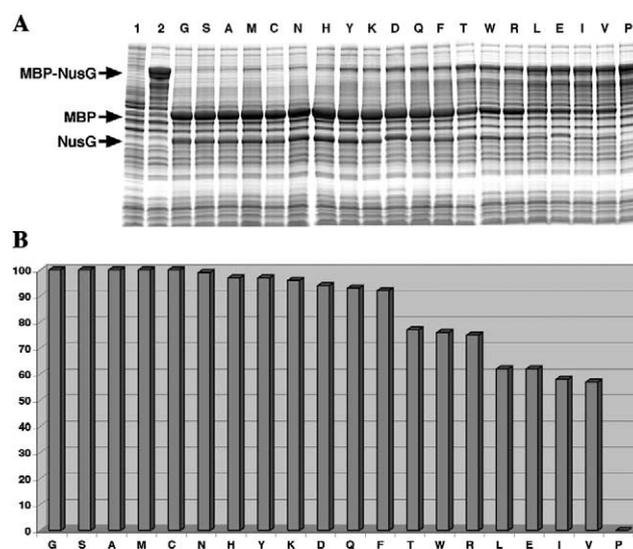


Fig. 2. Intracellular processing of MBP–NusG fusion proteins by TEV protease. (A) Samples of the total intracellular protein from *E. coli* DH5 α Z1 cells producing both TEV protease and one of twenty otherwise identical MBP–NusG proteins with different amino acid residues in the P1' position (indicated in single letter code) were prepared as described (Materials and methods) and analyzed by SDS-PAGE (10–20% Tris–glycine gradient gels). Lane 1: DH5 α Z1 cells containing no expression vector. Lane 2: DH5 α Z1 cells containing the TEV protease expression vector pRK603 and the MBP–NusG fusion protein vector pKM631 (with Gly in P1') induced with IPTG only. The positions of the intact fusion protein (MBP–NusG) and the digestion products (MBP, NusG) are indicated by arrows. (B) The percentage of each fusion protein that was cleaved *in vivo* by TEV protease was estimated from data obtained by laser scanning densitometry of Gelcode Blue (Pierce) stained gels, as described (Materials and methods).

that the requirement for a Ser or Gly in the P1' position is not as stringent as is generally presumed. In fact, quite a few of the substrates were processed to the same degree as the fusion proteins containing the canonical Gly or Ser in the P1' position. These experiments also revealed that the fusion proteins with β -branched hydrophobic residues or Glu in the P1' position are among the least efficient substrates for TEV protease.

In vitro processing of fusion proteins

Because many of the fusion proteins were processed nearly to completion by TEV protease *in vivo*, it was not possible to establish a hierarchy for the P1' specificity of the enzyme solely on the basis of the intracellular processing experiments. Therefore, taking advantage of the C-terminal hexahistidine tag, all 20 fusion proteins were purified by immobilized metal chelate affinity chromatography for *in vitro* processing experiments, as exemplified in Fig. 3 for the fusion protein with Gly in the P1' position. Wild-type TEV protease readily undergoes autolysis at a specific site to generate a truncated product with greatly diminished activity [22,24]. Accordingly, to ensure that all of the enzymes would remain active during the *in vitro* processing experiments, we employed a mutant protease (S219V) that is considerably more stable than the wild-type enzyme and just as catalytically active [22]. The fusion protein with Gly in the P1' position was used to establish the initial

conditions for *in vitro* processing because it possesses what is considered to be an optimal recognition site [25]. The substrate concentration was held constant at 1 mg/ml (14 μ M), while varying concentrations of TEV protease were utilized. Aliquots were removed from the reactions at regular intervals and analyzed by SDS-PAGE to monitor the extent of processing. At an enzyme to substrate ratio of 1:80, approximately 50% of the Gly variant was cleaved over the course of several hours (Fig. 4A).

It was clear from the intracellular processing experiments that not all of the fusion proteins would be cleaved as efficiently as the Gly variant. Therefore, only the substrates that were processed to at least 90% completion *in vivo* were examined *in vitro* at an enzyme to substrate ratio of 1:80 (Fig. 4A). These experiments revealed that the fusion proteins with Ser or Ala in the P1' position were processed even more readily than the Gly variant, which is the P1' residue most commonly employed in engineered TEV protease recognition sites. The Met, Cys, and His variants were also processed to a substantial extent after 3.5 h (30–40%).

Instead of increasing the duration of the reactions to obtain comparable data for the less efficient substrates, the enzyme concentration was adjusted while all of the other variables were held constant. At an enzyme to substrate ratio of 1:30, approximately 75% of the His variant was processed after 3.5 h (Fig. 4B). A substantial fraction (ca. 30–70%) of the Cys, Phe, Gln, Tyr, Asn, and Trp variants was also cleaved under these conditions but only a small portion of the remaining substrates was converted into products. Therefore, the processing of these less efficient substrates was subsequently compared at an enzyme to substrate ratio of 1:10. Under these conditions, the Asp, Thr, Glu, Leu, and Lys variants were all processed to a significant extent after 3.5 h (ca. 20–60%) but the remaining fusion proteins were not (Fig. 4C). Yet, except for the fusion protein with Pro in the P1' position, even the least efficient substrates for TEV protease (the Arg, Ile, and Val variants) were processed to a considerable extent (>50%) after only 3.5 h at an enzyme to substrate ratio of 1:2 (Fig. 4D). No processing (either specific or nonspecific) of the Pro variant was observed even after 24 h at an enzyme to substrate ratio of 2:1 (data not shown).

Processing of oligopeptide substrates

To corroborate the results obtained with the MBP–NusG fusion protein substrates, we also determined the kinetic parameters K_M and k_{cat} for a representative set of synthetic peptides with different residues in the P1' position (TENLYFQXGTRR–NH₂); these data are summarized in Table 1. Compared to the substrate with Gly in the P1' position, none of the P1' substitutions resulted

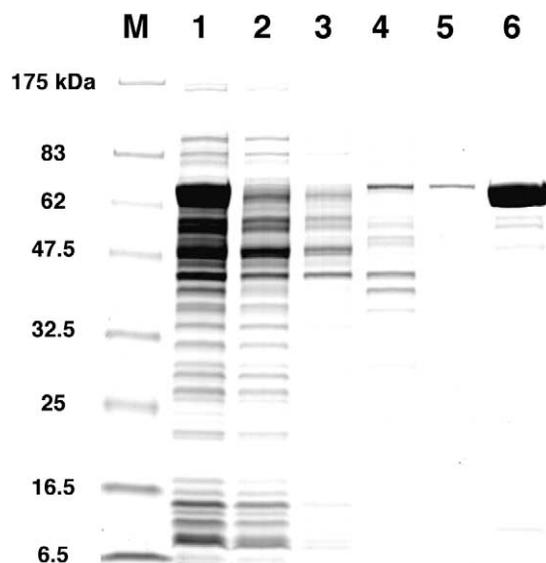


Fig. 3. Purification of MBP–NusG with Gly in the P1' position. Gel-code Blue stained SDS–polyacrylamide gel (10–20% Tris–glycine gradient) showing samples of MBP–NusG at various stages during purification. Lanes: M, broad-range molecular weight standards (kDa; kilodaltons); 1, soluble fraction of the total intracellular protein; 2, flow-through from the Ni–NTA column; 3–5, 20 mM imidazole washes; 6, protein eluted with 300 mM imidazole and dialyzed into TEV protease reaction buffer.

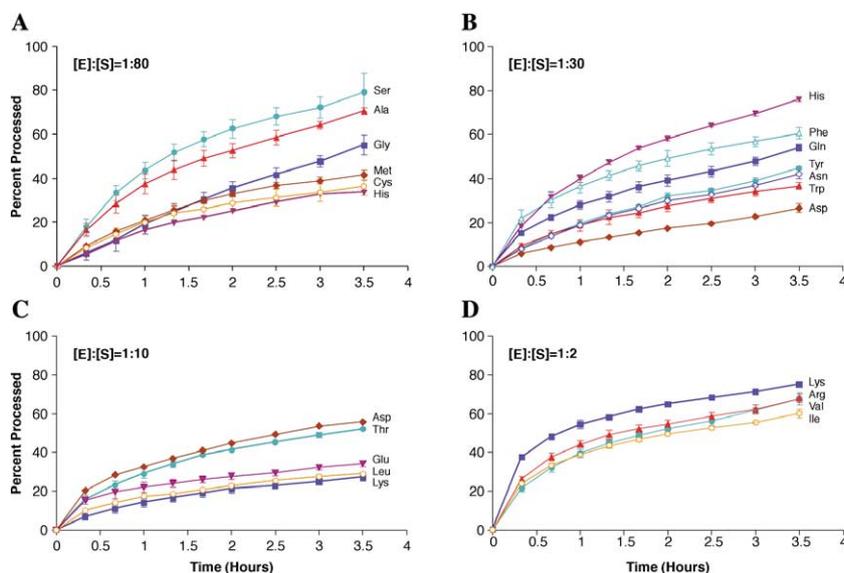


Fig. 4. In vitro processing of MBP–NusG fusion proteins by TEV protease. The concentration of the MBP–NusG substrates was $14\ \mu\text{M}$ ($1\ \text{mg/ml}$), and the concentration of TEV protease was either $0.18\ \mu\text{M}$ (A), $0.47\ \mu\text{M}$ (B), $1.4\ \mu\text{M}$ (C), or $7\ \mu\text{M}$ (D). Aliquots were removed at 30 min intervals over 3.5 h. The percentage of substrate cleaved by TEV protease at each time point (% cleaved) was estimated by densitometric analysis of Gelcode Blue stained SDS–polyacrylamide gels as described (Materials and methods). Each data point represents the average of three measurements with standard errors indicated by bars. The data for each substrate are identified by the amino acid residue in the P1' position of the TEV protease cleavage site (e.g., Gly).

Table 1
Kinetic parameters for oligopeptide substrates^a with amino acid substitutions in the P1' position

P1' residue	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M ($\text{mM}^{-1}\ \text{s}^{-1}$)
Ser	0.043 ± 0.006	0.194 ± 0.007	4.51 ± 0.65
Gly	0.087 ± 0.017	0.268 ± 0.025	3.08 ± 0.67
Ala	0.090 ± 0.015	0.271 ± 0.019	3.01 ± 0.54
Met	0.076 ± 0.007	0.180 ± 0.005	2.37 ± 0.23
Tyr	0.050 ± 0.010	0.031 ± 0.001	0.62 ± 0.13
Asp	0.057 ± 0.004	0.021 ± 0.001	0.37 ± 0.03
Gln	0.321 ± 0.025	0.073 ± 0.001	0.23 ± 0.02
Glu	0.094 ± 0.022	0.014 ± 0.001	0.15 ± 0.04
Lys	0.224 ± 0.046	0.027 ± 0.002	0.12 ± 0.03
Leu	0.240 ± 0.047	0.014 ± 0.001	0.06 ± 0.01
Val	N.D. ^b	N.D. ^b	0.010 ± 0.001

^a TENLYFQ(Xaa)GTRR–NH₂.

^b Not determined.

in an improved catalytic constant (k_{cat}), while both higher and lower K_M values were obtained. In general, the rank order of catalytic efficiencies (k_{cat}/K_M) of the peptides was in very good agreement with the hierarchy of processing efficiencies established for the fusion protein substrates, although the MBP–NusG fusion proteins with Gln and Trp in the P1' position seemed to be processed with comparatively greater efficiency than were the corresponding peptides. The minor discrepancies between the results obtained with the MBP–NusG fusion proteins and the peptide substrates could be due to the different assay conditions (e.g., pH, ionic strength), which may influence the kinetics of substrate cleavage in the case of some P1' residues.

Discussion

In addition to their obvious utility for protein purification, affinity tags can improve the yield of recombinant proteins, protect them from intracellular proteolysis, and in the case of MBP, enhance their solubility [1,2,4,5,7]. However, it is ordinarily desirable to remove the affinity tag from the passenger protein for functional and structural studies. Enzymatic methods are most commonly employed to remove affinity tags, yet not all proteases perform this task equally well. While factor Xa, enteropeptidase, and thrombin frequently cleave proteins at noncanonical sites [10,11], TEV protease is highly specific and active over a wide

range of pH and ionic strength [25]. However, until now it has been presumed that TEV protease exhibits a strong preference for Ser or Gly in the P1' position of its substrates. While it is certainly true that some of the P1' variants (e.g., the β -branched hydrophobic residues) are comparatively inefficient substrates for the enzyme, it is equally clear that many different residues can be accommodated in the P1' position of an otherwise canonical TEV protease recognition site with little impact on the efficiency of processing. These include Met and Ala, two of the most common natural N-terminal residues, and Cys. Polypeptides with an N-terminal Cys can be used to assemble segmentally labeled proteins for multidimensional heteronuclear NMR experiments [26], and the digestion of appropriate fusion proteins with TEV protease should provide a useful avenue for the creation of substrates for in vitro peptide ligation. Moreover, with the exception of the Pro variant, even the least efficient MBP–NusG fusion protein substrates were processed to a considerable degree after only a few hours when enough TEV protease was added to the reaction. Unlike factor Xa, thrombin, or enteropeptidase, TEV protease can be used at a high concentration without triggering nonspecific proteolysis. Consequently, our results imply that it should be possible to produce recombinant proteins with any N-terminal amino acid other than Pro by digesting a fusion protein with TEV protease, provided that the canonical recognition site is processed efficiently in the same context.

A previous mutational analysis of a naturally occurring TEV protease recognition site also led to the conclusion that certain amino acids other than Gly or Ser could occupy the P1' position without abolishing processing [14]. However, the only amino acids examined in this study were Ser, Ile, Asn, Arg, Thr, Phe, Cys, and Asp. Curiously, in contrast to the results reported here, these investigators observed that processing occurred even more rapidly when either Ile or Asn occupied the P1' position of the substrate than when Ser did, although neither of these residues is found in the P1' position of any natural potyviral cleavage sites. The other P1' variants that they tested were observed to be far less efficient substrates for TEV protease. To account for the discrepancies between our results, it is important to understand the differences between our methods. Because their substrate consisted of a relatively long segment of the TEV polyprotein, it is possible that tertiary interactions within the substrate or between the protease and the substrate influenced the results. Moreover, the protease they utilized was a nuclear inclusion body preparation that was composed of an equimolar mixture of the full-length 49 kDa nuclear inclusion protease and the 54 kDa viral replicase, whereas the TEV protease employed in this study was the soluble 27 kDa catalytic domain of the 49 kDa protease. In the previous study, nuclear inclusion bodies containing the 49 kDa TEV

protease were added to the products of an in vitro translation reaction containing the viral precursor. The concentration of protease in these reactions was approximately 2 μ M. The substrate concentration was not determined. However, considering the range of yields that have been obtained in the rabbit reticulocyte system [27], it could not have been more than 0.6 μ M and is likely to have been much lower than this. Thus, in the previous study the concentration of enzyme exceeded that of the substrate, which was also far below the K_M . The present study was conducted under more realistic reaction conditions, using pure preparations of enzyme and substrate with a substantial molar excess of the latter in most experiments.

Unfortunately, it is not possible to glean any direct information about the substrate-binding pocket of TEV protease because its three-dimensional structure has yet to be determined. However, our results indicate that TEV protease can accommodate a variety of amino acids in the P1' position of the cleavage site with relatively little impact on the efficiency of processing. The most efficient substrates tended to be those with the shortest side chains (e.g., Ser, Ala, and Gly), which are also the three residues that occur most frequently in the P1' position of natural cleavage sites for potyviral proteases. Thr is the next most common P1' residue in the natural substrates, but our results indicate that Met, Lys, His, and Asp are all tolerated as well or better than Thr. It therefore seems likely that some other kind of selective pressure restricts the range of residue types that occur in the P1' positions of the natural cleavage sites in potyviral polyproteins.

It has not escaped our notice that some of the NusG proteins generated by intracellular processing should be substrates for the N-end rule degradation pathway in bacteria [28]. In particular, N-terminal Tyr, Trp, Leu, and Phe are believed to be primary destabilizing residues in *E. coli*. Although no effort was made to directly measure the half-lives of NusG proteins with potentially destabilizing N-terminal residues, the fact that none of them appeared to be present in substoichiometric quantities relative to the free MBP generated by TEV protease digestion suggests that they were not rapidly degraded in vivo. Although the rank order of relative destabilizing activities among the twenty amino acids is thought to be invariant from one protein reporter to another in a given environment, the actual in vivo half-lives can differ greatly among different proteins bearing the same N-terminal residue [29]. Accordingly, determinants other than the N-terminal residue, such as the extreme thermostability of *Aquifex aeolicus* NusG, may make it an intrinsically poor substrate for the N-end protease ClpAP. In any case, we note that intracellular processing by TEV protease of fusion proteins with noncanonical residues in the P1' position could be used to generate N-end substrates in *E. coli* for further study.

Acknowledgments

We thank Karen Routzahn for constructing pKM631. A portion of the research reported in the document was originally published in a master's degree thesis by R.B.K. sponsored by the Department of Biomedical Science and submitted to the Graduate School of Hood College in Frederick, Maryland.

References

- [1] C.V. Maina, P.D. Riggs, A.G. Grandea, B.E. Slatko, L.S. Moran, J.A. Tagliamonte, L.A. McReynolds, C.D. Guan, An *Escherichia coli* vector to express and purify foreign proteins by fusion to and separation from maltose-binding protein, *Gene* 74 (1988) 365–373.
- [2] D.B. Smith, K.S. Johnson, Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase, *Gene* 67 (1988) 31–40.
- [3] A. Martinez, P.M. Knappskog, S. Olafsdottir, A.P. Doskeland, H.G. Eiken, R.M. Svebak, M. Bozzini, J. Apold, T. Flatmark, Expression of recombinant human phenylalanine hydroxylase as fusion protein in *Escherichia coli* circumvents proteolytic degradation by host cell proteases. Isolation and characterization of the wild-type enzyme, *Biochem. J.* 306 (1995) 589–597.
- [4] A. Jacquet, V. Daminet, M. Haumont, L. Garcia, S. Chaudoir, A. Bollen, R. Biemans, Expression of a recombinant *Toxoplasma gondii* ROP2 fragment as a fusion protein in bacteria circumvents insolubility and proteolytic degradation, *Protein Expr. Purif.* 17 (1999) 392–400.
- [5] T.R. Butt, S. Jonnalagadda, B.P. Monia, E.J. Sternberg, J.A. Marsh, J.M. Stadel, D. Ecker, S.T. Crooke, Ubiquitin fusion augments the yield of cloned gene products in *Escherichia coli*, *Proc. Natl. Acad. Sci. USA* 86 (1989) 2540–2544.
- [6] K.D. Pryor, B. Leiting, High-level expression of soluble protein in *Escherichia coli* using a His6-tag and maltose-binding protein double-affinity fusion system, *Protein Expr. Purif.* 10 (1997) 309–319.
- [7] R.B. Kapust, D.S. Waugh, *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused, *Protein Sci.* 8 (1999) 1668–1674.
- [8] C.W. Dykes, A.B. Bookless, B.A. Coomber, S.A. Noble, D.C. Humber, A.N. Hobden, Expression of atrial natriuretic factor as a cleavable fusion protein with chloramphenicol acetyltransferase in *Escherichia coli*, *Eur. J. Biochem.* 174 (1988) 411–416.
- [9] P.L. Rodriguez, L. Carrasco, Improved factor Xa cleavage of fusion proteins containing maltose binding protein, *Biotechniques* 18 (1995) 238–243.
- [10] G. Forsberg, B. Baastrup, H. Rondahl, E. Holmgren, G. Pohl, M. Hartmanis, M. Lake, An evaluation of different enzymatic cleavage methods for recombinant fusion proteins, applied on des(1-3)insulin-like growth factor I, *J. Protein Chem.* 11 (1992) 201–211.
- [11] R.C. Stevens, Design of high-throughput methods of protein production for structural biology, *Structure Fold. Des.* 8 (2000) R177–185.
- [12] C. Lauritzen, E. Tuchsén, P.E. Hansen, O. Skovgaard, BPTI and N-terminal extended analogues generated by factor Xa cleavage and cathepsin C trimming of a fusion protein expressed in *Escherichia coli*, *Protein Expr. Purif.* 2 (1991) 372–378.
- [13] W. Dieryck, V. Lullien-Pellerin, D. Marion, P. Joudrier, M.F. Gautier, Purification and activity of a wheat 9-kDa lipid transfer protein expressed in *Escherichia coli* as a fusion with the maltose binding protein, *Protein Expr. Purif.* 6 (1995) 597–603.
- [14] W.G. Dougherty, J.C. Carrington, S.M. Cary, T.D. Parks, Biochemical and mutational analysis of a plant virus polyprotein cleavage site, *EMBO J.* 7 (1988) 1281–1287.
- [15] W.G. Dougherty, S.M. Cary, T.D. Parks, Molecular genetic analysis of a plant virus polyprotein cleavage site: a model, *Virology* 171 (1989) 356–364.
- [16] R.B. Kapust, D.S. Waugh, Controlled intracellular processing of fusion proteins by TEV protease, *Protein Expr. Purif.* 19 (2000) 312–318.
- [17] A.G. Evdokimov, D.E. Anderson, K.M. Routzahn, D.S. Waugh, Overproduction, purification, crystallization and preliminary X-ray diffraction analysis of YopM, an essential virulence factor extruded by the plague bacterium *Yersinia pestis*, *Acta Crystallogr. D56* (2000) 1676–1679.
- [18] A.G. Evdokimov, J.E. Tropea, K.M. Routzahn, D.S. Waugh, Three-dimensional structure of the type III secretion chaperone SycE from *Yersinia pestis*, *Acta Crystallogr. D58* (2002) 398–406.
- [19] S.N. Ho, H.D. Hunt, R.M. Horton, J.K. Pullen, L.R. Pease, Site-directed mutagenesis by overlap extension using the polymerase chain reaction, *Gene* 77 (1989) 51–59.
- [20] J.H. Miller, *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1972.
- [21] J. Sambrook, E.F. Fritsch, T. Maniatis, *Molecular Cloning: A Laboratory Manual*, second ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.
- [22] R.B. Kapust, J. Tözsér, J.D. Fox, D.E. Anderson, S. Cherry, T.D. Copeland, D.S. Waugh, Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency, *Protein Eng.* 14 (2001) 993–1000.
- [23] P. Boross, P. Bagossi, T.D. Copeland, S. Oroszlan, L.M. Louis, J. Tözsér, Effect of substrate residues on the P2' preference of retroviral proteinases, *Eur. J. Biochem.* 264 (1999) 921–929.
- [24] T.D. Parks, E.D. Howard, T.J. Wolpert, D.J. Arp, W.G. Dougherty, Expression and purification of a recombinant tobacco etch virus NIa proteinase: biochemical analyses of the full-length and a naturally occurring truncated proteinase form, *Virology* 210 (1995) 194–201.
- [25] T.D. Parks, K.K. Leuther, E.D. Howard, S.A. Johnston, W.G. Dougherty, Release of proteins and peptides from fusion proteins using a recombinant plant virus proteinase, *Anal. Biochem.* 216 (1994) 413–417.
- [26] R. Xu, B. Ayers, D. Cowburn, T.W. Muir, Chemical ligation of folded recombinant proteins: segmental isotopic labeling of domains for NMR studies, *Proc. Natl. Acad. Sci. USA* 96 (1999) 338–393.
- [27] L. Jermutus, L.A. Ryabova, A. Plückthun, Recent advances in producing and selecting functional proteins by using cell-free translation, *Curr. Opin. Biotechnol.* 9 (1998) 534–548.
- [28] J.W. Tobias, T.E. Shrader, G. Rocap, A. Varshavsky, The N-end rule in bacteria, *Science* 254 (1991) 1374–1377.
- [29] A. Bachmaier, D. Finley, A. Varshavsky, In vivo half-life of a protein is a function of its amino-terminal residue, *Science* 234 (1986) 179–186.